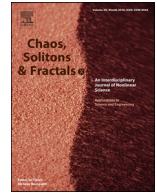




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



The Cyclicity of coronavirus cases: “Waves” and the “weekend effect”

Vladislav Soukhovolsky^a, Anton Kovalev^a, Anne Pitt^a, Katerina Shulman^b, Olga Tarasova^c, Boris Kessel^d

^a Federal Research Center, Siberian Branch of Russian Academy of Sciences, Krasnoyarsk, Russia

^b Carmel Medical Center, affiliated with Rappaport Medical School, Technion, Haifa, Israel

^c Siberian Federal University, Krasnoyarsk, Russia

^d Hillel Yaffe Medical Center, affiliated with Rappaport Medical School, Technion, Haifa, Israel

ARTICLE INFO

Article history:

Received 26 December 2020

Revised 19 January 2021

Accepted 20 January 2021

Available online 28 January 2021

Keywords:

Coronavirus

dynamics

time series

waves

week end effect

spectral analysis

autoregression

ABSTRACT

Introduction: Medical statistics is one of the “milestones” of current medical systems. It is the foundation for many protocols, including medical care systems, government recommendations, epidemic planning, etc. At this time of global COVID-19, credible data on epidemic spread can help governments make better decisions. This study’s aim is to evaluate the cyclicity in the number of daily diagnosed coronavirus patients, thus allowing governments to plan how to allocate their resources more effectively.

Methods: To assess this cycle, we consider the time series of the first and second differences in the number of registered patients in different countries. The spectral densities of the time series are calculated, and the frequencies and amplitudes of the maximum spectral peaks are estimated.

Results: It is shown that two types of cycles can be distinguished in the time series of the case numbers. Cyclical fluctuations of the first type are characterized by periods from 100 to 300 days. Cyclical fluctuations of the second type are characterized by a period of about seven days. For different countries, the phases of the seven-day fluctuations coincide. It is assumed that cyclical fluctuations of the second type are associated with the weekly cycle of population activity.

Conclusions: These characteristics of cyclical fluctuations in cases can be used to predict the incidence rate.

© 2021 Elsevier Ltd. All rights reserved.

Introduction

The possible cyclical nature of the new disease, with incidence “waves”, was predicted at the onset of the coronavirus pandemic, by analogy with the influenza epidemic [1,9]. However, the absence of long series of data, the occurrence of previous single-wave epidemics [12,13], and reliance on the classical mathematical model of the epidemic, which considered one “wave” of the disease [11], did not encourage quantifying the risk of epidemic “waves”. In addition, the biological factors causing the cyclical nature of epidemics remained unclear. For influenza, weather conditions are considered a modifying factor in the epidemic’s development; cases increase in autumn and winter. By analogy, it was assumed that the coronavirus epidemic would slow or even end by the summer of 2020. However, this did not happen, and since March 2020, sufficiently long series of incidence rates have been gathered in different countries to allow the cyclicity of coronavirus epidemics to be assessed and to identify the factors responsible for the disease’s recurrence. In this work, we studied the

cyclical nature of coronavirus incidence in different countries using the methods of spectral analysis of time series.

Objects and Methods

The database of the Johns Hopkins University (USA) was used for analysis [7]. The database for all countries contains the number $n(t)$ of cases accumulated since the epidemic’s beginning in a country from the date t_0 of the epidemic’s start to the day t_{max} (here, it is 12.08.2020). Using these data, we calculated the first relative differences, $p(t)/N = [n(t) - n(t-1)]/N$, characterizing the relative number of cases on a specific day t (N is a country’s population). By analyzing these series, one can characterize their general properties, such as the presence or absence of a time trend and cyclicity. Fig. 1 shows the time series of the relative number of daily reported cases in the United States from the disease’s onset to 12.08.2020.

We will use spectral analysis methods to assess periodic processes during the epidemic, according to which the time series is represented as the sum of sinusoids with a certain frequency

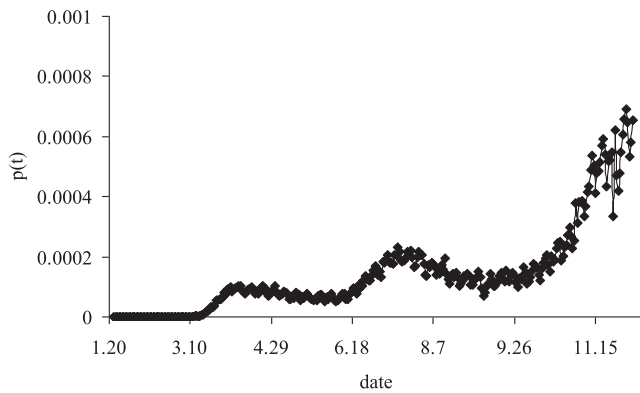


Fig. 1. The relative number $p(t)$ of daily registered patients with coronavirus in the United States from the onset of the disease to 12.08.2020.

and amplitude [2–4]. In the spectrum of the time series, the X-axis characterizes the frequency of oscillations with a certain frequency, and the Y-axis characterizes the spectral density - the amplitude $A(f)$ of oscillations with a frequency f . To calculate the spectra, the so-called Fourier transform is used [8,10,14,15]. To estimate the spectra of time series, the studied time series must be stationary in time, that is, their average values cannot change [18]. If the series is characterized by the presence of some monotonic trend (as can be seen in the US data), then it is necessary to "clear" the trend from the time series and then calculate it. For these time series, the ADF-test was used to estimate the stationarity of the series [5,6,17]. If the series turned out to be non-stationary, then to isolate the trend, the linear regression function $q = a + bt$ was calculated with respect to the value $p(t)$, or the series $w(t) = p(t) - q(t)$. If the significance of the coefficient of determination R^2 of the regression was low, the logarithmic transformation $p(t)$ was carried out, and for the series $\ln p(t)$, the logarithmically linear regression equation $q_1(t)$ was calculated and then the series $w_1(t) = \ln p(t) - q_1(t)$. If the time series $w(t)$ or $w_1(t)$ was stationary according to the ADF-test, its spectrum was calculated. If the series $w(t)$ still remained unsteady, the trend detection procedure could be repeated (however, this was never needed and the calculations were limited to a single trend detection procedure).

In addition, a procedure was used to isolate a number of second differences $\Delta w(t+1) = w(t+1) - w(t)$. If the number of first differences can be interpreted as the rate of the epidemic's increase or decrease, then a number of second differences can be represented as the acceleration of the epidemic's development. If the average value of time series is zero and the sum of the terms of the autocorrelation function of the series Δw is finite, then such a series will be stationary. After a transformation of the time series, their spectra were calculated. Below are the results of calculating the spectra for a number of countries.

Results

A log-transformed and detrending daily incidence series for the US is shown in Fig. 2.

There are three low-frequency "waves" of the disease for the USA during the period from the epidemic's beginning to early December 2020 (Fig. 2). After performing the Fourier transformation, the spectrum $s(f)$ of the time series was calculated (Fig. 3).

The peaks of the spectrum f_1, f_2, \dots determine the characteristic frequencies (or reciprocal values - the periods of fluctuations in the level of daily cases). As can be seen from Fig. 3, the maximum frequency $L_1 = 1/f_1$ for the US is 133 days. In addition to the peak at the frequency $f_1 = 0.0075$, the spectrum for the US is characterized by the presence of a peak at the frequency $f_2 = 0.143$

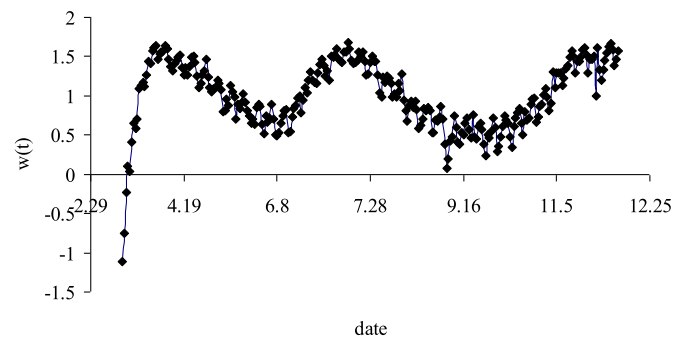


Fig. 2. Log-transformed and detrending daily incidence series for the United States.

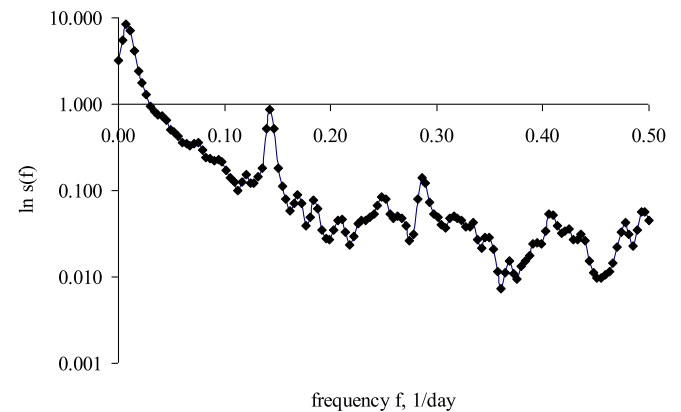


Fig. 3. Logarithmic spectrum of the number of first differences in the daily number of coronavirus cases in the United States.

Table 1

Wavelengths L_1 and L_2 and amplitudes A of high-frequency peaks in the spectrum of time series of coronavirus incidence for various countries (data sorted by the value $n(t_{\max})$ of the relative proportion of cases during the entire epidemic).

Country	L_1 , days	L_2 , days	$A(L_2) \cdot 10^7$	$n(t_{\max})$
Belgium	94	6.878	2.209	0.052
USA	133	6.957	0.093	0.046
Israel	143	6.952	1.056	0.038
Spain	320	6.957	3.245	0.036
Austria	288	7.024	0.555	0.035
France	320	3.441	8.240	0.034
Brazil	280	7.000	0.847	0.031
Italy	292	6.952	0.213	0.029
Chile	290	7.073	0.385	0.029
Sweden	286	6.976	4.590	0.029
United Kingdom	302	7.048	0.038	0.026
Russia	141	7.050	0.001	0.017
Denmark	286	6.810	0.016	0.016
Belarus	274	2.915	0.009	0.016
Germany	288	7.024	0.259	0.015
Iran	276	7.000	0.002	0.013
Saudi Arabia	282	3.032	0.002	0.010
India	280	6.880	0.002	0.007
Japan	101	6.809	0.00034	0.001
Egypt	278	11.120	0.00002	0.001
Korea, South	160	6.957	0.00007	0.001
China	-	3.951	0.00002	0.0001

(wavelength $L_2 = 7$ days). Spectra with similar peak frequencies are typical for the time series of the relative number of cases per day for other countries. Table 1 shows the wavelengths L_1 and L_2 of the incidence of coronavirus for a number of countries.

The duration ($1/\text{frequency}$) of epidemic waves differs significantly in different countries. Two groups of countries can be distinguished - countries with a short (from 94 to 160 days) dura-

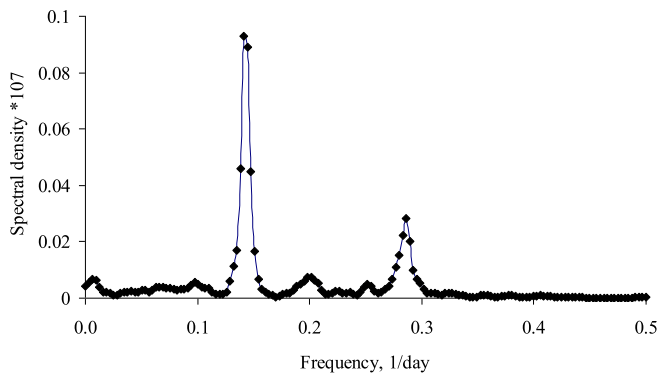


Fig. 4. Spectral density of second difference time series of cases in US.

tion of low-frequency fluctuations in incidence, and countries with a long duration of incidence (from 274 to 320 days). The twofold differences in the low-frequency incidence rate and the complete absence of waves for the incidence data in China indicate that the cause of the cyclical fluctuations is probably not related to the intrinsic properties of the virus itself, but is most likely determined by a country's characteristics, although it is difficult to find commonalities in the climates and the cultures, government response/action, behavior of population of Belgium, Japan, USA, Israel and Russia.

Most countries in Table 1 are characterized by fluctuations with a period of about 7 days. The transformation of the series of the first differences into the series of the second differences leads to the filtration of low-frequency components, and in the series of the second differences, the lowest frequency peaks are characterized by frequencies of about $f_{we} = 0.143$ 1/day ($L2 \approx 7$ days). The spectrum of second differences time series in incidence for the United States is shown in Fig. 4.

The $L2$ wavelengths in the series of second differences in this range are also given in Table 1, and the wavelength is close to 7 days for most of the analyzed data. The exceptions are France, Saudi Arabia, Belarus and China (the frequency of fluctuations is close to 3 days) and Egypt, with a frequency of 11 days. But the incidence statistics data for France turn out to be incorrect, because there are the errors in database and very often the number of accumulated cases $n(t)$ on data t is less than the value $n(t-1)$ on the previous day ($t-1$) in database.

Discussion

The analysis showed that the power of the peaks of the spectrum of the second differences was associated with indicators of case intensity—the total relative number of cases from the epidemic's beginning to 12.08.2020 (Fig. 5).

The reason for the peaks on the spectrums of the second difference time series for the number of sick patients remains unclear. There are no obvious biological causes for the disease's cyclicity, so the seven-day cyclicity may be associated with social effects. As a hypothesis, it could result from the so-called "weekend effect", when the patients with initial disease symptoms go to medical care providers only after the weekend, at the beginning of the next week. Fig. 6 shows trend-cleared and smoothed data on the distribution of the number of registered patients in the United States (weekends on Saturdays and Sundays) and Israel (weekends on Fridays and Saturdays).

In the US, at the beginning of the week, an increased number of patients need medical care because those who became ill on Saturday-Sunday will be registered as ill on Monday-Tuesday. In Israel, the number of registered patients is also at a minimum on

Table 2

Distribution of incidence by days of the week.

day of week	Brazil	Jordan	Turkey	Russia	Indonesia
Monday	-5.063	-0.474	-0.005	-0.065	-0.069
Tuesday	-0.063	0.350	0.014	-0.059	-0.035
Wednesday	3.424	0.980	0.016	0.007	0.036
Thursday	3.942	0.947	0.026	0.026	0.089
Friday	2.713	0.178	0.031	0.009	0.016
Saturday	-0.006	-1.017	-0.040	-0.014	0.003
Sunday	-4.948	-0.965	-0.043	-0.030	-0.040

weekends - on Friday and Saturday. A similar incidence distribution by days of the week is observed in other countries (Table 2).

However, there is no pronounced peak at the frequency $f_2 \approx 0.143$ in the spectrum of the second differences for a number of countries - Belarus, Saudi Arabia, China (about 3 days) Egypt (11 days). The lack of a weekend effect may be due to the lack of a complete weekend (in Egypt, only Friday is off) or to problems with the reliability of disease case registration in Belarus.

If the spectrum's peak in the transformed time series of cases is caused by a week-end effect, then there should be a shift in the phase between the transformed time series for countries whose weekends fall on Saturday - Sunday, the transformed time series for countries that have only Friday as a day off (i.e. Muslim countries), and Israel, which has weekends on Friday and Saturday.

However, the calculations showed that the spectrum's peaks do not appear for a number of countries. In particular, such peaks are not typical for the epidemic in China, and the spectrum for China in the first approximation can be seen as flat, with a lack of significant peaks in the spectrum and apparent frequency.

Perhaps the absence of significant peaks in the spectrum of the second differences time series of cases in China is associated with a more organized control of the disease, the state of the population and rapid testing for coronavirus. In China, the disease incidence is recorded as soon as it appears, and there is no seven-day cyclicity.

Earlier, an AR model was proposed to describe the dynamics of disease [16].

$$p(i) = a_0 + \sum_{j=1}^k a_j p(i-j) \quad (2)$$

where k is the autoregression order, a_0, \dots, a_k are the coefficients.

To estimate the order of autoregression according to the time series $\{p(i)\}$, the partial autocorrelation function (PACF) was calculated and the maximum significant value of $PACF(k)$ was determined. Further, Eq. (2) was considered as a linear regression equation with respect to the unknown values a_0, a_k , and these coefficients were calculated using standard methods for the known value of the autoregression order k .

The spectrum $I(f)$ of model (2) can be represented as the Fourier cosine transform of the autocorrelation function, calculated from the values of the coefficients a_j [2]:

$$I(f) = 2 \left(a_0 + 2 \sum_{j=1}^k a_j \cos 2\pi f j \right) \quad 0 \leq f \leq 0.5 \quad (3)$$

The maximum of $I(f_{\max})$ at a certain value of f_{\max} will characterize the wavelength $L1 = 1/f_{\max}$ of the incidence time series spectrum for a particular country. From the values of the autoregression order k and the coefficients a_j of Eq. (2), one can understand the reasons for the differences in the wavelengths $L1$ in the time series of the incidence of countries with sufficiently small values of $L1 \approx 140$ days (Israel, Russia) and countries with large values of $L1 \approx 280$ days (Brazil, Sweden). The parameters of model (2) for

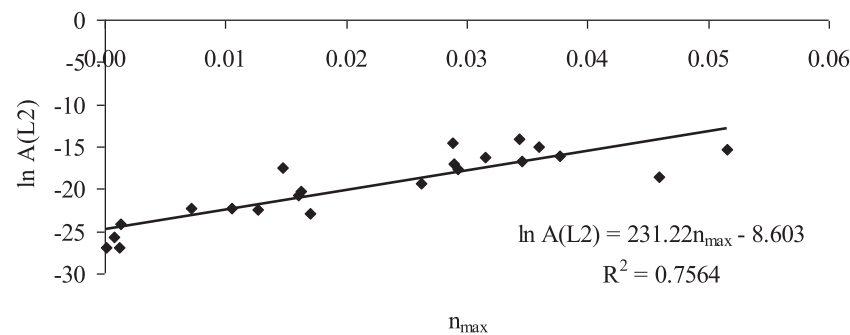


Fig. 5. Relationship between the relative number of cases from the beginning of the epidemic to 12.08.2020 and the intensity of the peak at frequency f_2 .

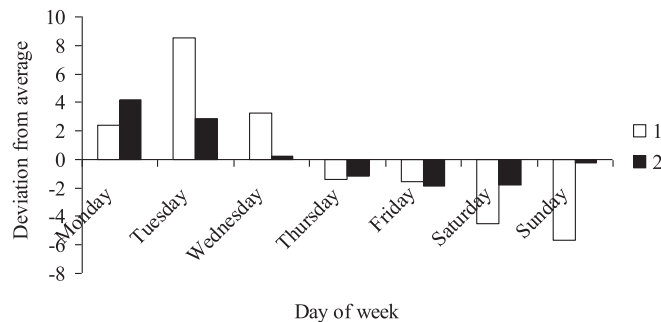


Fig. 6. Deviations from the trend-cleared and smoothed number of registered patients. 1 - US; 2 - Israel.

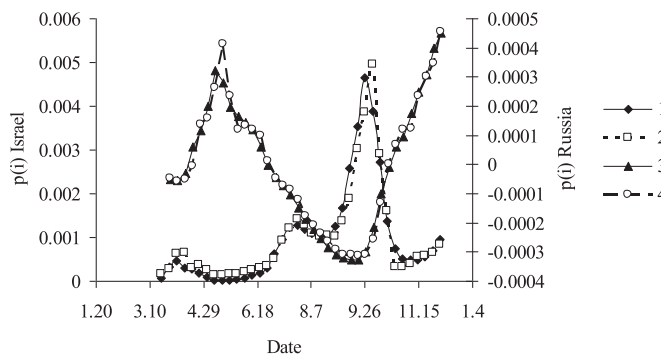


Fig. 7. Weekly time series (1) and model (2) for Israel and detrended weekly time series (3) and model (4) for Russia.

the weekly series of Israel, Russia, Brasil and Sweden are given in Table 3.

Time series of weekly data and models for Israel and Russia are shown in Fig. 7.

As shown in Table 3, all coefficients for the variables $p(j)$ are significant at the 0.05 level. Thus, we can conclude that the cyclicality of the coronavirus and, accordingly, the wavelength $L1$ of the time series of incidence is determined by the inertia k of the system. Perhaps inertia depends on the rate of withdrawal of sick patients from the population and from the number of asymptomatic patients. However, the weekly frequency is not very surprising; it exists for other medical conditions. The multiple epidemic waves were described in 1918 influenza pandemic. Yu D. et al, in their classic work, based on previous and his own improved modelling, demonstrated reactive social distancing, temperature, and school term could explain the observed multiple waves and final epidemic size [19]. Another evidence of multiple-waves outbreaks was recently published by Bo Xu. Possible explanations for the $L1$, could be seasonal climate effects, different peaking time in the North and

Table 3

Parameters of autoregressive equations for Israel, Russia, Brazil and Sweden. p-value.

Variables	Coefficients	Std. Err	t	
Israel				
a_0	0.0002	0.0001	2.1973	0.0351
$p(i-2)$	-0.7418	0.1147	-6.4667	0.0000
$p(i-1)$	1.5912	0.1156	13.7585	0.0000
K	2			
R^2	0.93			
F	210.6			
Russia				
a_0	0.0000	0.0000	0.5378	0.5942
$p(i-2)$	-0.8291	0.1137	-7.2909	0.0000
$p(i-1)$	1.7797	0.1071	16.6124	0.0000
K	2			
R^2	0.974			
F	625.7			
Brazil				
a_0	0.792	0.380	2.085	0.039
$p(i-7)$	0.535	0.078	6.883	0.000
$p(i-6)$	0.321	0.073	4.401	0.000
$p(i-1)$	0.296	0.078	3.770	0.000
k	7			
R^2	0.90			
F	161.6			
Sweden				
a_0	1.715	0.593	2.892	0.004
$p(i-5)$	-0.300	0.106	-2.823	0.005
$p(i-4)$	0.314	0.136	2.309	0.022
$p(i-3)$	-0.325	0.131	-2.477	0.014
$p(i-2)$	-0.315	0.123	-2.555	0.011
$p(i-1)$	1.358	0.072	18.987	0.000
k	5			
R^2	0.90			
F	322.7			

South hemispheres, and public behavior reaction to the pandemic [20]. The effect of global transportation, which remains unclearly limited, may also play the important role in this phenomenon.

What can be expected in the more distant future if mass vaccinations are carried out? Apparently, despite a decrease in the disease intensity, the seven-day cycle associated with the weekly rhythm of population activity may persist. The low-frequency cycle may also continue, especially if vaccinating most of the population takes a long time and if the quarantine is not continued, after vaccination begins.

Conclusions

The presented analysis demonstrates that at present time world widely, the morbidity dynamics is characterized by periodic low-frequency oscillations with wavelengths from 100 to 300 days and

of high-frequency oscillations with a period of about 7 days. An autoregressive model can be used to describe such fluctuations with high level of accuracy.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Vladislav Soukhovolsky: Conceptualization, Methodology, Software, Investigation, Validation, Writing - review & editing. **Anton Kovalev:** Conceptualization, Methodology, Software, Investigation, Validation, Writing - review & editing. **Anne Pitt:** Conceptualization, Methodology, Software, Investigation, Validation, Writing - review & editing. **Katerina Shulman:** Methodology, Investigation, Validation, Writing - review & editing. **Olga Tarasova:** Conceptualization, Methodology, Software, Investigation, Validation, Writing - review & editing. **Boris Kessel:** Conceptualization, Methodology, Software, Investigation, Validation, Writing - review & editing.

References

- [1] Alessa A, Faezipour M. A review of influenza detection and prediction through social networking sites// Theoretical Biology. Medical Modelling 2018;15(1):2–29.
- [2] Anderson TW. Statistical Analysis of Time Series. N.Y.: Wiley; 1971. p. 704.
- [3] Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time Series Analysis: Forecasting and Control. N.Y.: Wiley; 2015. p. 712.
- [4] Brockwell PJ, Davis RA. Introduction to Time Series and Forecasting (Springer Texts in Statistics). Berlin: Springer; 2016. p. 439.
- [5] Hamming R.W. Digital filters. Courier Corporation. 1989. 284 p.
- [6] Hayashi F. Econometrics. Princeton and Oxford. Princeton University Press; 2000.
- [7] <https://github.com/CSSEGISandData/COVID-19> Data Repository by Johns Hopkins (Johns Hopkins University, 2020).
- [8] Jenkins GM, Watts DG. Spectral analysis and its applications. San Francisco: Holden-Day; 1969. p. 525.
- [9] Keeling MJ, Rohani P. Modeling Infectious Diseases in Humans and animals. New Jersey: Princeton University Press; 2007. p. 408.
- [10] Kendall M.G., Stewart A. The advanced theory of statistics: design and analysis, and time series. L.: C. Griffin. 1973. 780 p.
- [11] Kermack WO, McKendrick AG. In: A Contribution to the Mathematical Theory of Epidemics // Proceedings of the, 115. Royal Society; 1927. p. 700–21.
- [12] Legrand J, Grais RF, Boelle PY, Valleron AJ, Flahault A. Understanding the dynamics of Ebola epidemics // Epidemiology and Infection 2007;135(04):610–21 Iss.
- [13] Li ZL, Zhang LM. Mathematical Model of SARS Prediction and Its Research Progress. Journal of Mathematical Medicine 2004;17(6):481–4.
- [14] Marple S. L., Jr. Digital spectral analysis: with applications. Englewood Cliffs, N.J.: Prentice-Hall. 1987. 492 p.
- [15] Shumway, R. H. Time Series Analysis and Its Applications with R Examples / R. H. Shumway, D. S. Stoffer. – New York: Springer Science Business Media, LLC, 2006. – 202 p.
- [16] Soukhovolsky V., Kovalev A., Pitt A., Kessel B. A new modelling of the COVID 19 pandemic// Chaos, Solitons & Fractals (IF 3.764) Pub Date: 2020-06-20, DOI: 10.1016/j.chaos.2020.110039
- [17] Stock JH, Watson MW. Introduction to Econometrics. Addison-Wesley; 2011. p. 832.
- [18] Wei W. Time series analysis: univariate and multivariate methods. Boston: Pearson. 2006. 614 p.
- [19] Yu D, Lin Q, Chiu AP, He D. Effects of reactive social distancing on the 1918 influenza pandemic. PLoS One 2017;12(7):e0180545.
- [20] Xu B, Cai J, He D, Chowell G, Xu B. Mechanistic modelling of multiple waves in an influenza epidemic or pandemic. J Theor Biol 2020 Feb 7;486:110070.